Madelyn Rojas Castro, Marina Travanca, Marta Fernandez Avalos, David Valentin Conesa, Emmanuel Lagarde

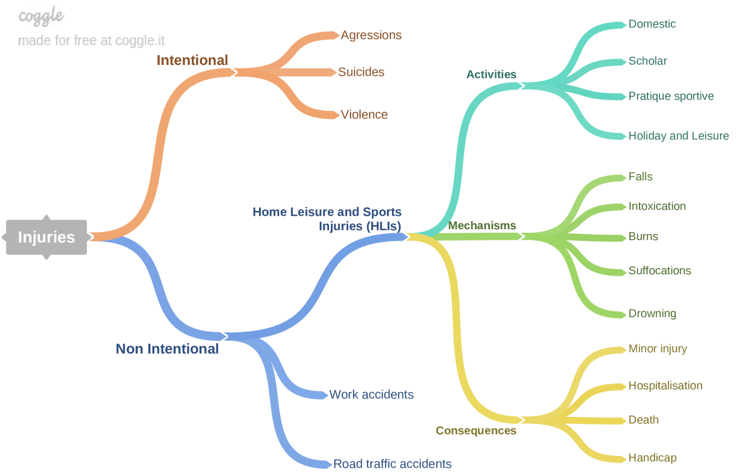
Usefulness of Bayesian modeling in risk analysis and prevention of *Home Leisure and Sport Injuries* (HLIs)

Madelyn Rojas Castro[1],
Marina Travanca[1], Marta Avalos[1], David Conesa[2], Emmanuel Lagarde[1]

[1] Bordeaux Population Health Research Center INSERM U1219,
Injury epidemiology, transport, occupation (IETO).
[2] Departamento de Investigación Operativa, Universidad de Valencia

18 January 2018

Home Leisure and Sport Injuries (HLIs)



Injuries Epidemiological Context

World Health Organization (WHO)

- **Injuries** are the 4th cause of mortality in the EU.
- **230 thousand** annual deaths (EU), 58 % are HLIs.

In France :

- **HLIs are the 3th cause of mortality**.
- **Leading cause of childhood mortality**.

Each year:

- **20 thousand** deaths, 5 times more than traffic accidents.
- **5 million** emergencies.
- **11 million** injuries

L' Observatoire MAVIE

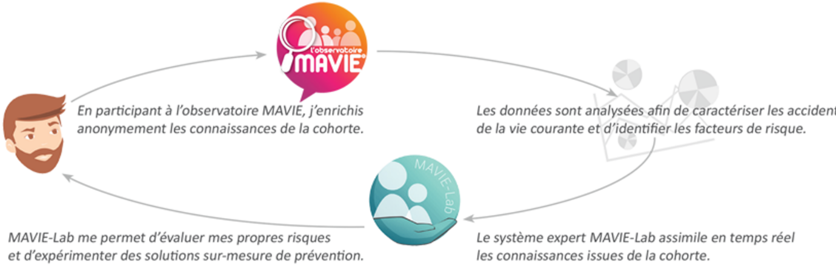
- Prospective online cohort study of *HLIs*
- Currently, MAVIE has more than **26 thousand** volunteers in France during 3 years of recruitment (target sample **100 thousand** volunteers).

Objectives

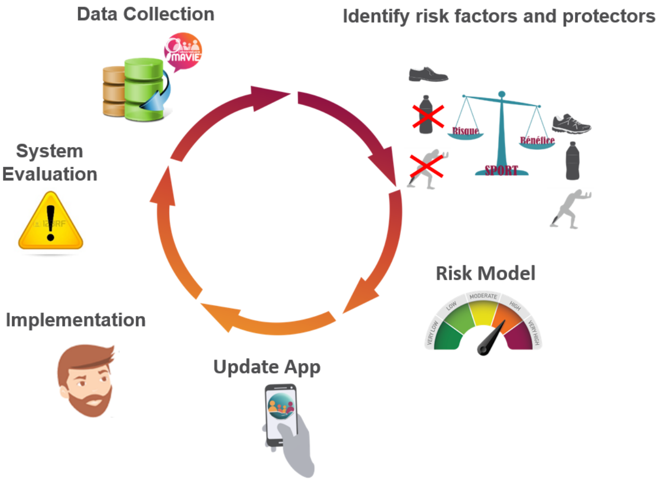
- Identify the **risk factors** associated with the *HLIs* occurrence and severity.
- Implement **prevention measures** to reduce the number of victims.

MAVIE-Lab (mHealth)

- **Mobile app** including a **DSS (Decision Support System)**
- To self-management of HLIs risks (Evaluation).
- To experience personalized **prevention solutions** to reduce the risk of injury (Mitigation).



MAVIE-Lab Development



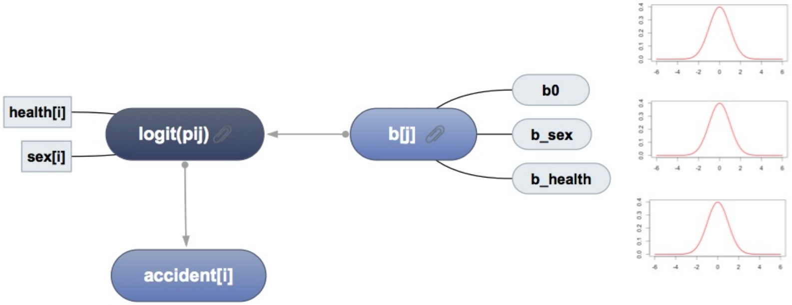
Modeling Problems MAVIE data

1. Reduced number of injuries declared by each activity.
2. Missing values.
3. Under-representation between injuries reported and those occurred.
4. Complex relationships between risk factors.

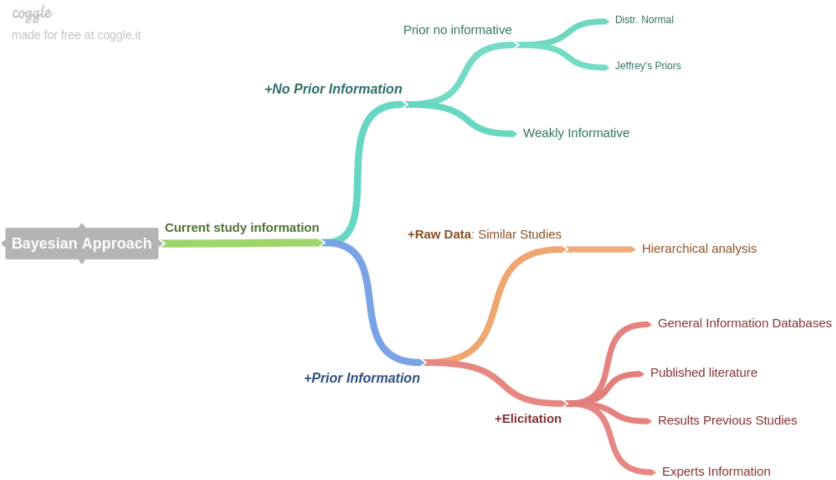
Bayesian Generalized Lineal Models (Logistic Regressión)

$$p(\boldsymbol{\theta} \mid \boldsymbol{X}, \boldsymbol{y}) \propto p(\boldsymbol{\theta})\, l(\boldsymbol{\beta}, \boldsymbol{\sigma} \mid \boldsymbol{X}, \boldsymbol{y})$$

$$posterior \propto prior \times likelihood$$



Bayesian Generalized Lineal Models (Logistic Regression)



Methodological Proposal

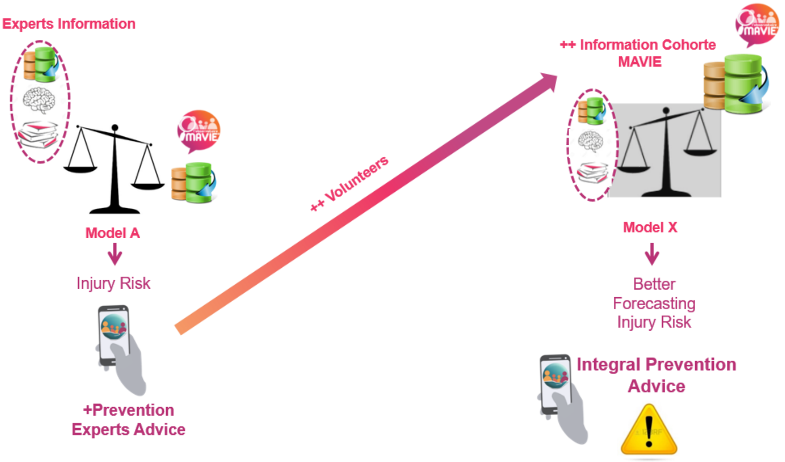
Priors experts elicitation

Reason for using Bayesian experts elicitation

1. The knowledge is limited or incomplete.
2. The evidence is inconsistent missing or ambiguous.
3. The questions are complex and the relations between variables.
4. To deal with bias and uncertainties.
5. To integrate different sources of knowledge.

(A. B. Knol *et al.* 2010)

Bayesian experts elicitation (MAVIE-Lab)



Exploratory analysis objectives

Principal Objective

To explore **Bayesian modeling methodologies** for being used in MAVIE-Lab development

- To explore *experts elicitation* to improve the estimation of model parameters.
- To explore the use of automatic *selection models methods* since the Bayesian approach.

Model Selection BMA *Bayesian Model Averaging*

The BMA is a weighted average of the posterior distributions for each parameter for all possible models.

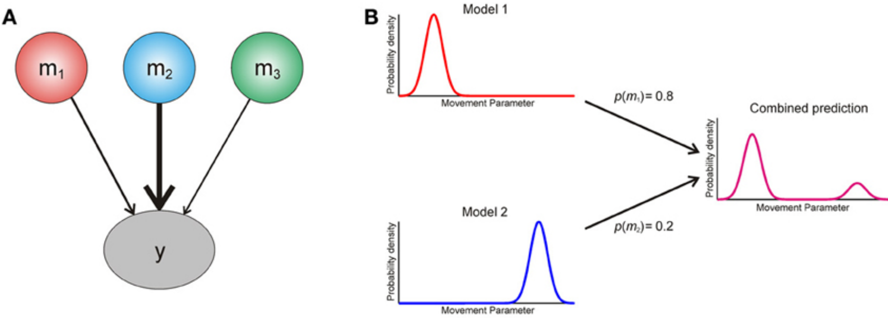


Figure taken from (FitzGerald *et al.* 2014).

Work Database

- **Sample of MAVIE Cohort:** N = 4,345 (March 2017). Volunteers over 18 years old, who had completed and validated the questionnaire .
- **Injuries:** 603 Reported Injuries (13.87 %).
- **Variables:** Explanatory variables: 20 categorical or categorized variables (Factors associated with HLIs occurrence).
  - **Demographic:** Age and sex.
  - **Previous Injuries**.
  - **Physical and mental health:** BMI, Health problems, depression, anxiety, hyperactivity, drowsiness and concentration.
  - **Consumption:** Medicines, alcohol, tobacco and cannabis.
  - **Sport Practice:** Sports, use of compression and maintenance accessories.

Methodology

MAVIE Data

↓

Multiple Imputation

↓

BY Logistic Model (Non informative *priors*)

↓

Model Selection (BMA *Bayesian Model Averaging)*

↓

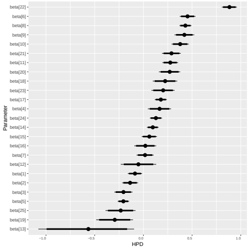
Elicitation Prior Distributions

↓

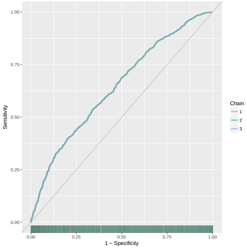
BY Logistic Model (Informative *priors*)

Results

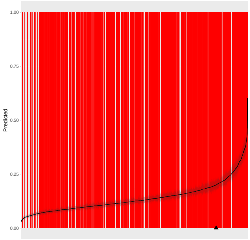
BY Logistic Model (non informative *priors*)



(a) CI 95%

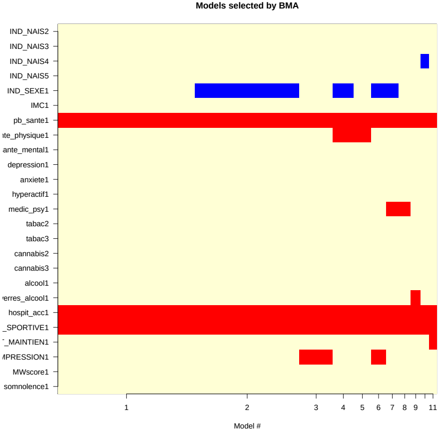


(b) ROC (AUC=0,58)



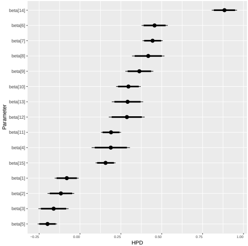
(c) Separation Graphic

Model Selection BMA: Better models according to BIC criteria

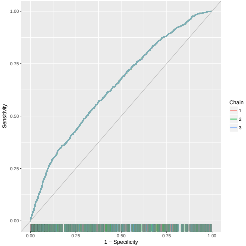


Selected Variables: Sex, Health Prob., Previous HLIs, Sports.

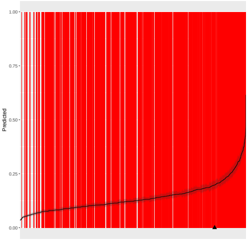
Model BMA Variables:
BY Logistic Model (Non informative *priors*)



(d) CI 95%



(e) ROC (AUC=0.64)

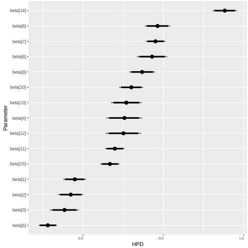


(f) Separation Graphic

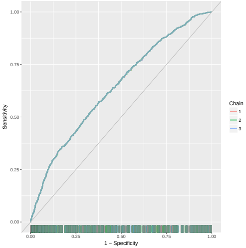
Informative Prior Distributions: *Elicited* values prior distributions models. Reference Study (Injuries in France) (Lefèvre & Mhiri 2015).

Variables	**OR**	**Mean β**	**Variance β**
Sex (F)	0.44	-0.36	0.04
Age 50-60	1.20	0.08	0.05
Age 40-50	1.40	0.15	0.04
Age 30-40	1.40	0.15	0.05
Age 18-30	**2.10**	0.32	0.04
Health Prob.	1.74	0.24	0.07
Smoker	1.64	0.21	0.03
Ex-smoker	1.34	0.13	0.03
Alcohol	**5.72**	0.76	0.28

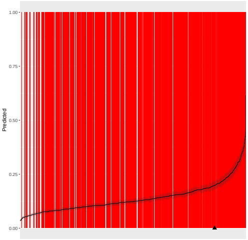
Informative Prior Distributions



(g) CI 95%



(h) ROC (AUC=0.6)



(i) Separation Graphic

Results BY model: (BMA, Prior Information)

	OR	Devest	2.5%	97.5%
Intercept	-	-	-	-
Age (50-60)	0.95	0.035	0.883	1.020
Age (40-50)	0.93	0.037	0.855	1.002
Age (30-40)	0.89	0.043	0.812	1.973
Age (18-30)	1.29	0.076	1.154	1.443
Sex (F)	0.80	0.024	0.760	0.851
Sport	**1.58**	0.065	1.467	1.719
Health Prob.	1.57	0.050	1.476	1.668
Smokers	1.54	0.074	1.397	1.688
Ex-smokers	1.44	0.060	1.331	1.567
Physic Health	1.35	0.052	1.252	1.453
Psc.Medic.	1.22	0.038	1.146	1.294
Mantenim. Acc.	1.29	0.072	1.150	1.434
Comp. Acc.	1.31	0.065	1.188	1.440
Previous HLIs	**2.41**	0.095	2.236	2.605
Alcohol	1.18	0.037	1.112	1.254

Discussion

- The model does not allow separating profiles of greater or lesser injury risk.
- The variables included *do not explain* the occurrence of injuries.

Bayesian approach remains appropriate for the development of MAVIE-Lab.

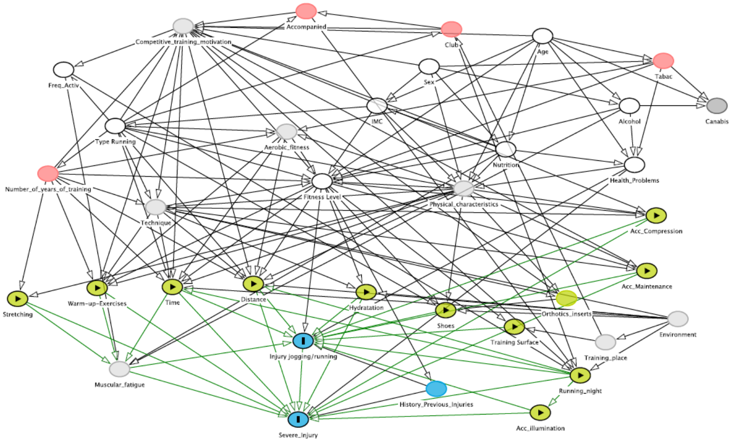
Solutions and ongoing work

1. To make **more specific models by type of injury**, including the most important variables in each case (Example: **Sport Injuries** *PhD Project*).

2. To perform a **formal elicitation** (Devilee & A. Knol 2011):
   - Experts selection
   - Uncertain evaluation
   - Elicitation protocol

3. To perform **Bayesian Network models** including besides the relationships between variables (*probabilistic* and *graphical modeling*) (Mujalli *et al.* 2016).

Variables relations in Running Injuries (DAG)



Conclusion

- In conclusion, **Bayesian statistic** and the expert's **elicitation** are powerful tools for the construction of **expert system** to be included in mHealth. This methodology makes possible to combine **statistical data** and experts information as for example **medical advice**.

Figure: **mHealth** (Figure taken from web-site UNC Gillinds School of Global Public Health)

Bibliography

Devilee, J. & Knol, A. *Software to support expert elicitation An exploratory study of existing software Software to support expert elicitation.* Tech. rep. (2011), 1–100.

FitzGerald, T. H. B., Dolan, R. J. & Friston, K. J. Model averaging, optimal inference, and habit formation. *Frontiers in Human Neuroscience* **8,** 457. ISSN: 1662-5161 (June 2014).

Knol, A. B., Slottje, P., Van Der Sluijs, J. P. & Lebret, E. The use of expert elicitation in environmental health impact assessment: a seven step procedure. *Environmental Health* **9.** doi:10.1186/1476-069X-9-19. http://www.ehjournal.net/content/9/1/19 (2010).

Lefèvre, B. & Mhiri, E. F. Facteurs sociodémographiques et pratiques associés aux accidents liés à la pratique physique et sportive. *Science et Sports* **30,** 126–133 (2015).

Mujalli, R. O., Lopez, G. & Garach, L. Bayes classifiers for imbalanced traffic accidents datasets. *Accident Analysis and Prevention.* doi:10.1016/j.aap.2015.12.003 (2016).

Usefulness of Bayesian modeling in risk analysis and prevention of *Home Leisure and Sport Injuries* (HLIs)

Madelyn Rojas Castro[1],
Marina Travanca[1], Marta Avalos[1], David Conesa[2], Emmanuel Lagarde[1]

[1] Bordeaux Population Health Research Center INSERM U1219,
Injury epidemiology, transport, occupation (IETO).
[2] Departamento de Investigación Operativa, Universidad de Valencia

18 January 2018